Systematic and Applied Microbiology 44 (2021) 126223

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.com/locate/syapm

Proposal of *Carbonactinosporaceae* fam. nov. within the class *Actinomycetia*. Reclassification of *Streptomyces thermoautotrophicus* as *Carbonactinospora thermoautotrophica* gen. nov., comb. nov

Camila Gazolla Volpiano [a,1], Fernando Hayashi Sant'Anna [b,1], Fábio Faria da Mota [c], Vartul Sangal [d], Iain Sutcliffe [d], Madhaiyan Munusamy [e], Venkatakrishnan Sivaraj Saravanan [f], Wah-Seng See-Too [g], Luciane Maria Pereira Passaglia [a], Alexandre Soares Rosado [h,i,*]

[a] *Departamento de Genética and Programa de Pós-graduação em Genética e Biologia Molecular, Instituto de Biociências, 9500, Bento Gonçalves Ave, Porto Alegre, RS, Brazil*
[b] *PROADI-SUS, Hospital Moinhos de Vento, 630, Ramiro Barcelos Porto Alegre, RS, Brazil*
[c] *Laboratório de Biologia Computacional e Sistemas, Instituto Oswaldo Cruz, 4365, Brasil Ave, Rio de Janeiro, RJ, Brazil*
[d] *Faculty of Health and Life Sciences, Northumbria University, Newcastle upon Tyne, United Kingdom*
[e] *Temasek Life Sciences Laboratory, 1 Research Link, National University of Singapore, Singapore 117604, Singapore*
[f] *Department of Microbiology, Indira Gandhi College of Arts and Science, Kathirkamam, Pondicherry, India*
[g] *Division of Genetics and Molecular Biology, Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia*
[h] *LEMM, Laboratory of Molecular Microbial Ecology, Institute of Microbiology Paulo de Góes, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil*
[i] *BESE, Biological and Environmental Sciences and Engineering Division, KAUST, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia*

ARTICLE INFO

Article history:
Received 6 April 2021
Revised 27 May 2021
Accepted 1 June 2021

Keywords:
Actinobacteria
Chemoautotrophic bacteria
Genomics metrics
Phylogenetic systematics

ABSTRACT

Streptomyces thermoautotrophicus UBT1^T has been suggested to merit generic status due to its phylogenetic placement and distinctive phenotypes among *Actinomycetia*. To evaluate whether '*S. thermoautotrophicus*' represents a higher taxonomic rank, '*S. thermoautotrophicus*' strains UBT1^T and H1 were compared to *Actinomycetia* using 16S rRNA gene sequences and comparative genome analyses. The UBT1^T and H1 genomes each contain at least two different 16S rRNA sequences, which are closely related to those of *Acidothermus cellulolyticus* (order *Acidothermales*). In multigene-based phylogenomic trees, UBT1^T and H1 typically formed a sister group to the *Streptosporangiales-Acidothermales* clade. The Average Amino Acid Identity, Percentage of Conserved Proteins, and whole-genome Average Nucleotide Identity (Alignment Fraction) values were ≤58.5%, ≤48%, ≤75.5% (0.3) between '*S. thermoautotrophicus*' and *Streptosporangiales* members, all below the respective thresholds for delineating genera. The values for genomics comparisons between strains UBT1^T and H1 with *Acidothermales*, as well as members of the genus *Streptomyces*, were even lower. A review of the '*S. thermoautotrophicus*' proteomic profiles and KEGG orthology demonstrated that UBT1^T and H1 present pronounced differences, both tested and predicted, in phenotypic and chemotaxonomic characteristics compared to its sister clades and *Streptomyces*. The distinct phylogenetic position and the combination of genotypic and phenotypic characteristics justify the proposal of *Carbonactinospora* gen. nov., with the type species *Carbonactinospora thermoautotrophica* comb. nov. (type strain UBT1^T, = DSM 100163^T = KCTC 49540^T) belonging to *Carbonactinosporaceae* fam. nov. within *Actinomycetia*.

Abbreviations: AAI, Average Amino acid Identity; AF, Alignment Fractions; ANI, Average Nucleotide Identity; ANIb, ANI using the BLASTn; BBHs, bidirectional best hits; CDSs, protein-coding sequences; dDDH, digital DDH; gANI, whole-genome ANI; GBDP, Genome BLAST Distance Phylogeny; HGT, horizontal gene transfer; KOs, KEGG (Kyoto Encyclopedia of Genes and Genomes) orthologous (KOs); LPSN, List of Prokaryotic names with Standing in Nomenclature; LTP, Living Tree Project; MiSI, Microbial Species Identifier; ML, Maximum Likelihood; NISEs, Non-Homologous Isofunctional Enzymes; POCP, Percentage of Conserved Proteins; TYGS, the Type (Strain) Genome Server.

* Corresponding author at: BESE, Biological and Environmental Sciences and Engineering Division, KAUST, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia.
E-mail address: alexandre.rosado@kaust.edu.sa (A.S. Rosado).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.syapm.2021.126223
0723-2020/

Introduction

The genus *Streptomyces* Waksman and Henrici 1943, belonging to order *Streptomycetales*, family *Streptomycetaceae* within the class *Actinomycetia* (former *Actinobacteria* [1]), is one of largest bacterial genera as currently defined, with more than 600 validly named species according to the List of Prokaryotic names with Standing in Nomenclature (https://lpsn.dsmz.de/genus/streptomyces). Members of the genus are highly significant because of their complex lifecycles, including sporulation, which have made them important as model organisms for studies of bacterial genetics and ecology [2–4]; and because members of the genus are highly proficient producers of secondary metabolites of biomedical and biotechnological interest, notably antibiotics and anticancer compounds [5–7]. Given the size of the genus *Streptomyces*, attempts have been made to determine internal structure of the genus through phylogenetic characterization of species groups, and its relationships to other taxa within the family *Streptomycetaceae* [8–11], although more comprehensive phylogenomic studies are required to resolve interspecies and suprageneric structure [6,11].

Genomic metrics have become the gold standard for defining taxonomic ranks, especially species designations among prokaryotes as they provide a reproducible, reliable, and a highly informative means to infer relatedness directly between genomes sequences [12,13]. In particular, average nucleotide identity (ANI) using the BLASTn algorithm (ANIb) and the Genome BLAST Distance Phylogeny (GBDP)-based digital DDH (dDDH) methods have been widely used to determine species boundaries and confirm identification [12,14,15]. The use of phylogenetic analyses in addition to Average Amino acid Identity (AAI), Percentage of Conserved Proteins (POCP), and whole-genome ANI (gANI) coupled with Alignment Fractions (AF) metrics have been also proposed to demarcate genus and higher taxa [13,16–18].

Streptomyces thermoautotrophicus Gadkari et al. 1991 has been suggested to merit generic status since it does not cluster within the *sensu stricto Streptomyces* clade in a tree inferred with the GBDP using formula *d5* [11] and was located apart from the clade containing members of the family *Streptomycetaceae* in a phylogeny from 14 well-conserved proteins [19]. This species was described by Gadkari et al. [20] based on characteristics of a single strain, UBT1^T, isolated from soil covering a charcoal burning pile. Strain UBT1^T is of interest as a sporulating aerobic thermophile, exhibiting growth at 40–68 °C, likely reflecting its isolation source. It was claimed to be a CO- and H_2-oxidizing obligate chemolithoautotrophic bacterium [20] and to produce a biochemically distinct, oxygen insensitive nitrogenase [21,22]. Later, MacKellar et al. [19] isolated a second '*S. thermoautotrophicus*' strain, H1, from another burning charcoal pile near an active coal seam fire. Multiple CO dehydrogenase gene clusters have been identified in the genomes of strains UBT1^T and H1; however, genes encoding nitrogenase enzymes seem to be absent. In addition, strains H1 and UBT1^T were unable to grow on Noble agar or to incorporate $^{15}N_2$ into biomass, besides growing heterotrophically on pyruvate. As a result, MacKellar et al. [19] proposed the reclassification of '*S. thermoautotrophicus*' as non-diazotrophic, facultative chemolithoautotrophic bacteria. Nevertheless, the chemolithoautotrophic metabolism of '*S. thermoautotrophicus*' distinguishes it from members of the genus *Streptomyces* [11]. In addition, the presence of eleven (UBT1^T) or nine (H1) biosynthetic gene clusters for secondary metabolites in the '*S. thermoautotrophicus*' genomes [19] is relatively low compared to other members of *Streptomyces* [5], reflecting the comparatively small genome sizes (~5 Mb) of these two strains. Finally, the circularity of the H1 genome distinguishes it from *Streptomyces sensu stricto*, where most genomes are linear [23].

Although earlier studies have presented convincing evidence that '*S. thermoautotrophicus*' should be reclassified into a novel genus [11,19], formal taxonomic proposals to achieve this have not been made due to concerns about the type strain availability and ambiguity concerning its suprageneric relationships. Here, we revisit the taxonomy of '*S. thermoautotrophicus*' and propose its reclassification as *Carbonactinospora thermoautotrophica* gen. nov., comb. nov., within the *Carbonactinosporaceae* fam. nov.

Material and methods

16S rRNA sequence identity analysis

The 16S rRNA gene sequence data for the type strains within the class *Actinomycetia* were retrieved from the SILVA SSU r138.1 database [24] (link to the full license: https://creativecommons.org/licenses/by/4.0/legalcode). The search criteria in the Living Tree Project (LTP) dataset were set as follows: "*Actinobacteria*" in the taxonomy field, sequence length >1400 nucleotides, sequence quality >90, and type strains (search term "[T]" in the strain field). Sequences were downloaded as an alignment in FASTA format containing gaps. As in the SILVA database, *Acidothermus cellulolyticus* is classified within the order *Frankiales* [25], it was manually corrected to *Acidothermales*.

The 16S rRNA genes sequences of strains '*S. thermoautotrophicus*' UBT1^T [20] and H1 [19] were extracted from the RefSeq genome assemblies GCF_001543895 and GCF_001543925, respectively, and subsequently aligned using SINA 1.2.11 [26]. A consensus alignment between the genomic '*S. thermoautotrophicus*' and the SILVA alignments was obtained. Positions containing gaps were removed from the alignment and an identity matrix for the resulting 858 positions was computed using Bioedit v. 7.0.5.3. The Python library Seaborn v. 0.11.1 was utilized for building boxplots of the16S rRNA identities for each order within the class *Actinomycetia*. Additionally, sequence identity was assessed by comparing the 16S rRNA sequences of the UBT1^T and H1 strains with the sequences from EzBioCloud [27], a quality-controlled 16S rRNA server database.

Phylogenetic analyses

Genera within the class *Actinomycetia* were identified in the lineage file available in https://github.com/zyxue/ncbitax2lin (v. 2019-02-20), which was generated from the NCBI taxonomy dump. Subsequently, the type species of each genus were retrieved according to the names provided on LPSN using the script "get_type_genus.py" (available at https://github.com/fhsantanna/bioinfo_scripts). All the available proteomes for *Actinomycetia* type species in the NCBI Assembly RefSeq database were downloaded for further analyses. The proteomes of the *Embleya* and *Trebonia* type species were later included manually since these genera were not available in the lineage file. Lastly, the proteome of UBT1^T and H1 were included in the final sequence dataset.

Two different approaches were carried out for the phylogenetic reconstruction based on the concatenated alignment of orthologous proteins. The first approach utilized the AMPHORA2 [28] pipeline for the identification of universal taxonomic markers in the *Actinomycetia* proteomes. For this purpose, the "phylogenomics-tools" scripts were utilized [29]. The markers *dnaG*, *infC*, *nusA*, *pgk*, *pyrG*, *rplK*, *rpoB*, *rpsC*, and *smpB* were excluded from the analyses as they were present in either multiple copies or at low representation among the type species. A total of 253 taxa

remained in the dataset after the exclusion of proteomes that did not present the final 22 markers (Supplementary Material). After, each marker protein was aligned using MUSCLE [30] v. 3.8.31 and concatenated. Positions containing gaps were excluded and the final 3184 amino acids alignment was utilized as input for the phylogenetic reconstruction based on the Maximum Likelihood (ML) method in the PhyML 3.0 server [31]. The substitution model was selected based on the Akaike Information Criterion to select LG + G + I as the best model, with an estimated gamma shape parameter of 0.829 and an estimated proportion of invariable sites of 0.165. Branch support was assessed using aLRT SH-like [32].

The second approach was a protein-based core genome phylogeny using a *de novo* identification of phylogenetic markers. Core ortholog groups of the previously selected strains were identified using bidirectional best hits (BBHs) algorithm implemented in GET_HOMOLOGUES [33] pipeline build 31072020, excluding inparalogs and using minimal blast searches. Once the core proteins were identified, GET_PHYLOMARKERS [34] v. 2.2.8.1 was used with default parameters (-R 1 -t PROT options) for finding optimal ortholog clusters for phylogenomic reconstruction. This approach is based on three main filters: exclusion of alignments containing recombinant sequences, removal of reconstructions that deviate from expectations of the multispecies coalescent, and elimination of poorly resolved gene trees. Top-scoring gene alignments were concatenated into a supermatrix, which was utilized to estimate the species-tree with the ML method. The phylogenetic trees were processed with Newick utilities [35], whose functionalities include taxa renaming and tree pruning (i.e. removing clades and only keeping those of interest).

In order to obtain a genome tree using GBDP, the genome sequence data were uploaded to TYGS, the Type (Strain) Genome Server [36]. In brief, the determination of closest type strain genomes was done in two complementary ways: first, the UBT1^T and H1 genomes were compared against all type strain genomes available in the TYGS database via the MASH algorithm, a fast approximation of intergenomic relatedness [37], and, then the ten type strains with the smallest MASH distances were chosen for each '*S. thermoautotrophicus*' genome. Second, an additional set of ten closely related type strains was determined via the 16S rDNA gene sequences. These were extracted from UBT1^T and H1 genomes using RNAmmer [38] and each sequence was subsequently BLAST searched [39] against the 16S rDNA gene sequence of each of the currently 13,011 type strains available in the TYGS database. This was used as a proxy to find the best 50 matching type strains (according to the bitscore) for each '*S. thermoautotrophicus*' genome and to subsequently calculate precise distances using the GBDP approach under the algorithm 'coverage' and distance formula *d5* [40]. For the calculation, local-alignment programs are used to align a genome *X* against a genome *Y*, and vice versa, producing a set of high-scoring segment pairs (HSPs). These matches are then transformed to a single distance value d(X, *Y*) by applying the formula *d5*, which is calculated as two times the sum of identical base pairs over all HSPs ($2 \times I_{xY}$) divided by the total length of all HSPs found in both genomes ($H_{xY} + H_{YX}$) [41,42], rescaled for phylogenetic inference and with branch support values based on resampling [43].These distances were finally used to determine the 10 closest type strain genomes for each of the '*S. thermoautotrophicus*' genomes. For the GBDP tree reconstruction, all pairwise comparisons among the set of genomes were conducted using GBDP under the algorithm 'trimming' and distance formula *d5*. The resulting distances were used to infer a balanced minimum evolution tree with branch support via FASTME 2.1.4 including SPR postprocessing [44]. Branch support was inferred from 100 pseudo-bootstrap replicates each. The trees were rooted at the midpoint [45] and visualized with PhyD3 [46].

Genome and proteomic metrics

Proteomic and genomic relatedness metrics were computed comparing '*S. thermoautotrophicus*' to type species from the order *Streptosporangiales*, and the genera *Acidothermus*, *Catenulispora*, *Frankia*, *Micromonospora*, *Pseudonocardia*, *Sporichthya* and *Streptomyces*.

gANI and AF values were obtained by the Microbial Species Identifier (MiSI) method using ANIcalculator 2014-127 v. 1.0 (https://ani.jgi.doe.gov/html/home.php?page=introduction). gANI is calculated for a pair of genomes by averaging the nucleotide identity of orthologous genes identified as BBHs, which are the genes that show ≥70% sequence identity and ≥70% alignment of the shorter gene. AF is calculated as a fraction of the sum of the lengths of BBH genes divided by the sum of the lengths of all genes in a genome [47].

POCP values were obtained with the script "POCP.sh" (available at https://figshare.com/articles/POCP_calculation_for_two_genomes/4577953/1), which was written based on Quin et al. [17]. For POCP calculation, the conserved proteins between a pair of genomes are determined by aligning all the protein sequences of a genome *X* against a protein's sequences from a genome *Y*, using the BLASTP aligner. Conserved proteins are defined as presenting a match with an $<1e^{-5}$ E value, > 40% of sequence identity, and >50% of an alignable region of the query protein sequence. The POCP (X, *Y*) % is calculated as $[(C_x + C_y)/(T_x + T_y)] \times 100$, where C represents the conserved number of proteins and T represents the total number of proteins on the respective genome.

AAI analyses were performed using the script "aai.rb" implemented in the Enveomics Collection [48]. For AAI calculation, the conserved genes between a pair of genomes are determined by aligning all protein-coding sequences (CDSs) of a genome *X* against a translated database of genome *Y*, using the TBLASTN aligner. Conserved CDSs are defined as presenting >30% of sequence identity at the amino acid level and >70% of an alignable region of the query CDS sequence. The matching segment from the genomic sequence is extracted and the reverse search with BLASTX is used to determine the presumably orthologous fraction of conserved genes between the two genomes (two-way BLAST). The two-way AAI (X, *Y*) % is measured by the average amino acid identity of all two-way BLAST conserved genes between the genomes, as computed by the BLAST algorithm [49]. For evaluating the AAI diversity between '*S. thermoautotrophicus*', *Streptomyces,* and *Streptomycetales* type strains, all available proteomes from these taxa were utilized for AAI computation as described above. As a control, *Streptomyces albus*, the type species of the genus *Streptomyces*, was compared to the same taxa.

Scatter plots showing the relationship between AAI and POCP and between AAI and AF were generated using the Python library Seaborn v. 0.11.1.

Taxonomic profiling of proteomes

AAI-profiler, which is a webserver dedicated to taxonomic identification [50], was employed to perform proteome-wide sequence searches using '*S. thermoautotrophicus*' UBT1^T and H1 genomes. AAI-profiler computes AAI between a query proteome and all target species in the UniProt database [50]. Each protein is binned considering the taxonomic attribution of the closest counterpart in the database. A taxonomic profile of the proteome of interest is built considering the counts of the target taxa, and these frequencies are weighted by the percent identity of the match to the query. Given that '*S. thermoautotrophicus*' is already included in the AAI-profiler database, the taxonomic profile excluded hits of the top-ranked taxon, which are from '*S. thermoautotrophicus*' itself.

KEGG orthologous analysis

To find metabolic divergences between '*S. thermoautotrophicus*' and other *Streptomyces* spp., the predicted amino acid sequences from the genomes of strains UBT1^T and H1 were compared to the KEGG (Kyoto Encyclopedia of Genes and Genomes) orthologous (KOs) belonging to the 71 *Streptomyces* spp. obtained from KEGG database [51] (Supplementary Material).

Results and discussion

Diversity of the 16S rRNA genes from 'S. thermoautotrophicus'

To evaluate the taxonomic position of '*S. thermoautotrophicus*' within *Actinomycetia*, we first conducted a 16S RNA gene identity sequence analysis of strains UBT1^T and H1. As previously reported [19], the genome of UBT1^T contains three 16S rRNA genes, two of which are identical to each other (locus tags TH66_RS04095 and TH66_RS03010) whilst the other is divergent (TH66_RS22860), presenting 94% identity to the other two. The genome assembly of H1 contains two 16S rRNA genes, one of which (LI90_RS08655) is identical to the TH66_RS04095/TH66_RS03010 pair, and the other (LI90_RS18525) is identical to the divergent copy TH66_RS22860.

The presence of multiple 16S RNA gene copies within a single bacterial genome has been observed before. Indeed, bacteria can harbour more than 20 copies of this marker gene [52]. The presence of intragenomic heterogeneity of 16S rRNA ≥6% was also reported in some thermophiles, such as the *Firmicutes* members *Desulfotomaculum kuznetsovii* DSM 6115^T and *Thermoanaerobacter tengcongensis* MB4^T, and the *Actinobacteria* member *Thermobispora bispora* DSM 43833^T [53,54]. This may constitute an ecological strategy [55–57] to adapt the bacterial cellular machinery to perform under different temperatures [58], with different copies being

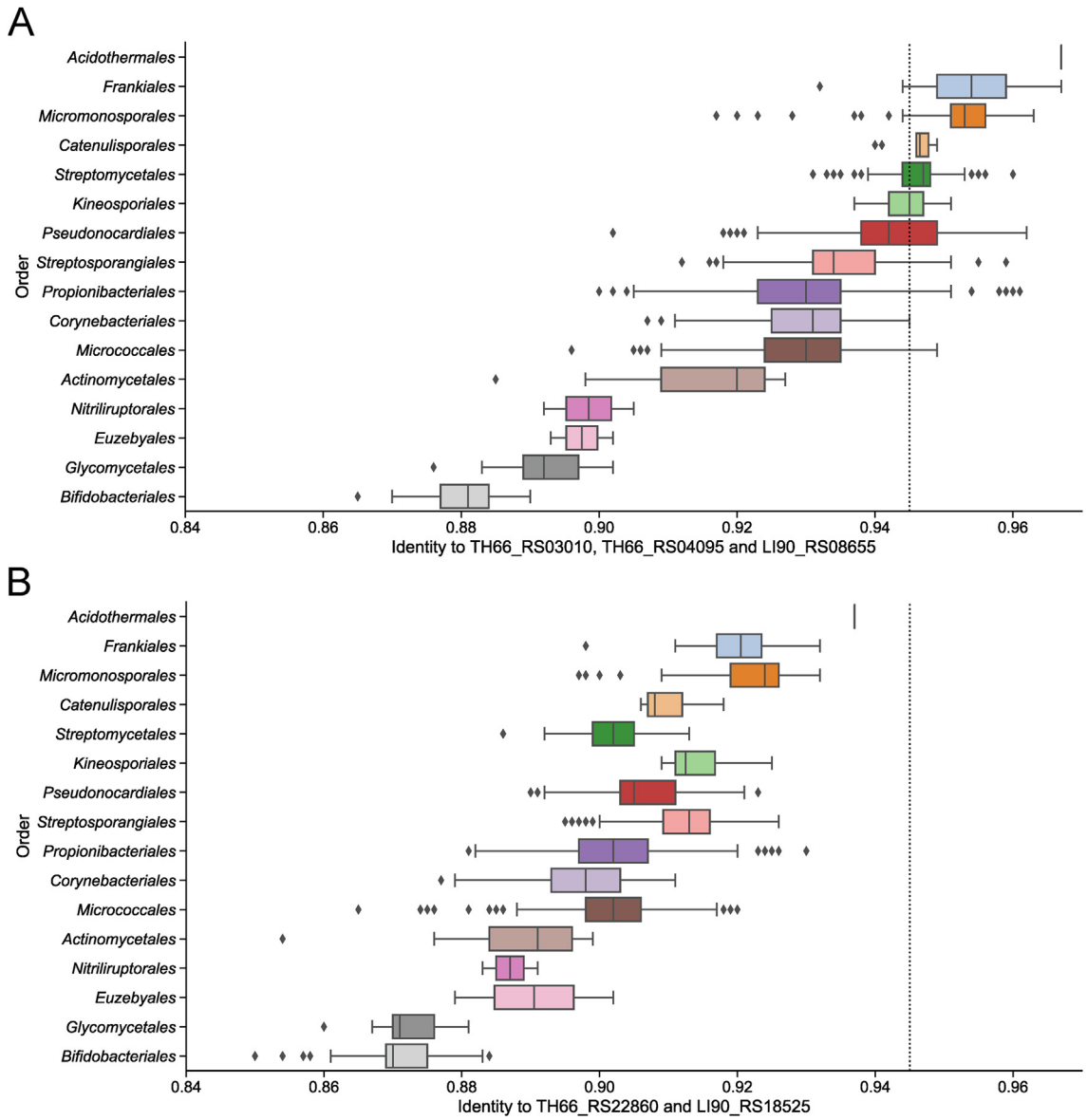

Fig. 1. Distribution of identity values of 16S RNA sequences of '*S. thermoautotrophicus*' UBT1^T and H1 to those of other *Actinomycetia* species summarized by order. (A) Identity values of the 16S rRNA alignments to the genes TH66_RS03010, TH66_RS04095 and LI90_RS08655. (B) Identity values of the 16S rRNA alignments to the genes TH66_RS22860 and LI90_RS18525. Boxplots show the three quartile values of the distribution. Horizontal bars inside the boxes represent the median (second quartile). Error bars extend to points that lie within 1.5 IQRs of the lower and upper quartile, and observations that are outside this range are displayed independently (outliers). Boxplots are colored according to the taxonomic Order. Vertical dashed lines represent the genus circumscription identity threshold of 94.5%.

functional under different environmental conditions [59]. In addition to the biases introduced from PCR [60,61], the presence of multiple different 16S rRNA gene copies is another strong argument against relying only on 16S rRNA gene phylogeny in species delineation in traditional polyphasic approach.

To identify 16S rRNA gene relatedness at the genus level, each copy from UBT1^T and H1 was compared to 2,792 16S rRNA sequences from type strains of *Actinomycetia* species available in the SILVA database. According to this analysis, TH66_RS04095/TH66_RS03010/LI90_RS08655exhibit identities above the 94.5%

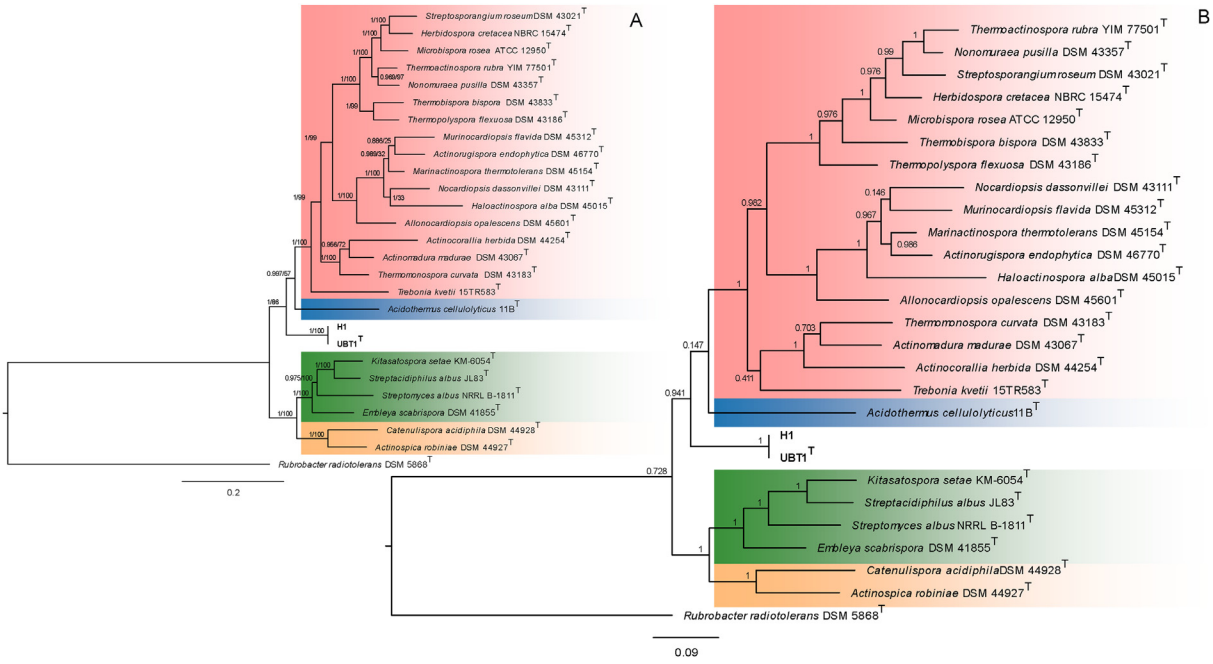

Fig. 2. Multiprotein phylogenies of *Actinomycetia* type-species and '*S. thermoautotrophicus*' UBT1^T and H1. (A) Phylogenetic reconstruction based on the core-proteome proteins of *Actinomycetia* type species using ML (IQ-TREE). Clade support values are shown next to the nodes. The first value corresponds to approximate Bayes branch support values and the second one to the UFBoot (Ultra Fast bootstrap) values. (B) Phylogenetic reconstruction of AMPHORA2 proteins using PhyML. aLRT values greater than 70% are shown next to the nodes. The trees are drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic trees. *Rubrobacter radiotolerans* DSM 5868^T is the outgroup. Only representative taxa are shown in these trees, which are the pared-down version of the complete trees (253 genomes) shown in Figs. S1–S3.

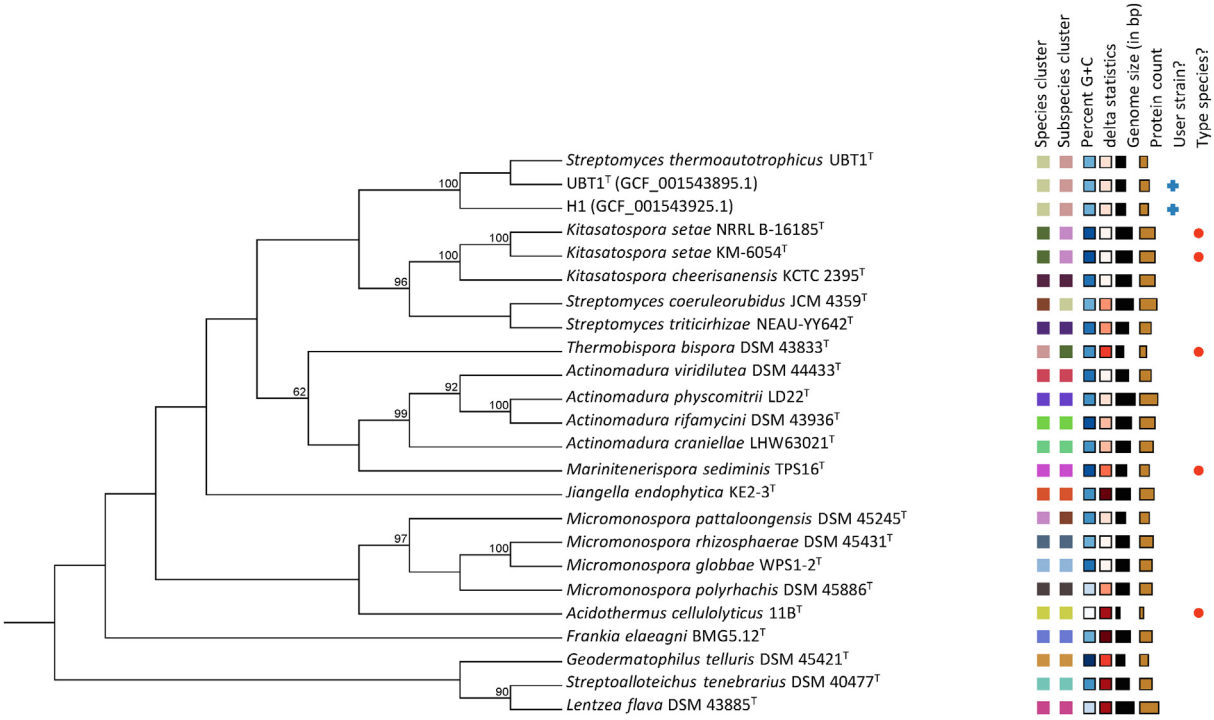

Fig. 3. Tree inferred with FastME 2.1.6.1 from GBDP distances calculated from '*S. thermoautotrophicus*' UBT1^T and H1 genome sequences. The branch lengths are scaled in terms of GBDP distance formula d_5. The numbers above branches are GBDP pseudo-bootstrap support values >60% from 100 replications, with an average branch support of 66%. The tree was rooted at the midpoint.

genus circumscription threshold [62] with 16S rRNA sequences from 87 non-*Streptomyces* and six *Streptomyces* type strains (not including the type species *Streptomyces albus*), being closely related to *A. cellulolyticus* from the order *Acidothermales* with 96.7% of identity (Fig. 1A). Sequences from representatives of *Acidothermales*, *Frankiales,* and *Micromonosporales* exhibit the highest identities to TH66_RS04095/TH66_RS03010/LI90_RS08655.

The more divergent TH66_RS22860/LI90_RS18525 copies did not belong to any recognized phylotypes at the genus level when compared with sequences from the *Actinomycetia* dataset (Fig. 1B), and even with the 65,797 entries in the EzBioCloud 16S rRNA database (Table S1). In both analyses, *A. cellulolyticus* stood out in presenting 93.7% identity to these divergent copies.

The current understanding of the evolutionary forces shaping the genomes of *Actinomycetia* is limited [63]; however, McDonald and Currie [64] analyzed 122 *Streptomyces* genomes and found that the acquisition and retention of genes through horizontal gene transfer (HGT) are surprisingly rare in this genus. Considering these findings, one of these 16S rRNA sequences can be assumed to be ancestral to strains UBT1^T and H1 while the other in each genome seems to be the product of a more recent duplication, rather than an HGT event.

Phylogenetic placement of 'S. thermoautotrophicus' within Actinomycetia

A multigene-based phylogenetic approach should be the choice for defining genera or higher taxa according to the minimal standards for the use of genome data for the taxonomy of prokaryotes [13]. Thus, to identify the current closest relatives to '*S. thermoautotrophicus*' UBT1^T and H1, two different approaches were employed to reconstruct the evolutionary history of UBT1^T, H1 and an additional set of 251 type species of *Actinomycetia* with genomes/proteomes available. We first reconstructed a ML phylogenetic tree with the concatenated protein sequences from 22 conserved single-copy genes identified in the assemblies with the AMPHORA2 pipeline. In addition, we performed a *de novo* approach for the identification of nine ortholog genes/ubiquitous proteins in the *Actinomycetia* type species genomes which were appropriate for phylogenomic analysis, with only three of them, encoding proteins of the 50S ribosomal subunit, also present in the AMPHORA2 dataset. Both phylogenetic reconstructions (Fig. 2 and Figs. S1–S3) infer that the genus *Streptomyces* does not form a clade with UBT1^T and H1, and the latter strains share a last common ancestor with *A. cellulolyticus* and members of the *Streptosporangiales* clade, thus belonging to a deeply branching lineage.

In the previous phylogenomic analysis that included '*S. thermoautotrophicus*', MacKellar et al. [19] highlighted the unusual position of the UBT1^T and H1 genomes as being closely related to *Acidothermus* and *Streptosporangiales* (*Streptosporangium, Thermobifida, Thermobispora,* and *Thermomonospora*), and distinct from the clade containing the families *Streptomycetaceae* and *Catenulisporaceae*. Therefore, the authors proposed that UBT1^T and H1 do not belong to the genus *Streptomyces* and instead are nearer to families including *Acidothermaceae* and *Streptosporangiaceae.* The proposal of a generic status for '*S. thermoautotrophicus*' was also supported by Nouioui et al. [11], in a tree inferred with GBDP formula *d5* [11], where UBT1^T branched away from core *Streptomyces* before *Kitasatospora* and *Streptacidiphilus*, forming a sister group to the core *Streptomyces-Kitasatospora-Streptacidiphilus* clade. The position inferred by Nouioui et al. [11], however, conflicts with MacKellar et al. [19] and our phylogenetic reconstructions based on ML estimations (Fig. 2), where UBT1^T forms a sister group with *Acidothermus* and members of *Streptosporangiales.*

To obtain a current GBDP tree for '*S. thermoautotrophicus*', the genome sequence data for UBT1^T and H1 were uploaded to TYGS (Fig. 3). The distance-based tree demonstrated that UBT1^T and H1 form a distinct group of *Actinomycetia*, however, due to the low branch support, the sister groups for the '*S. thermoautotrophicus*' strains could not be delimited precisely. The phenetics or distance-based approaches, such as GBDP, try to fit a tree to a matrix of pairwise genetic distances, therefore, reflecting the number of nucleotide or amino-acid substitutions [65]. In contrast,

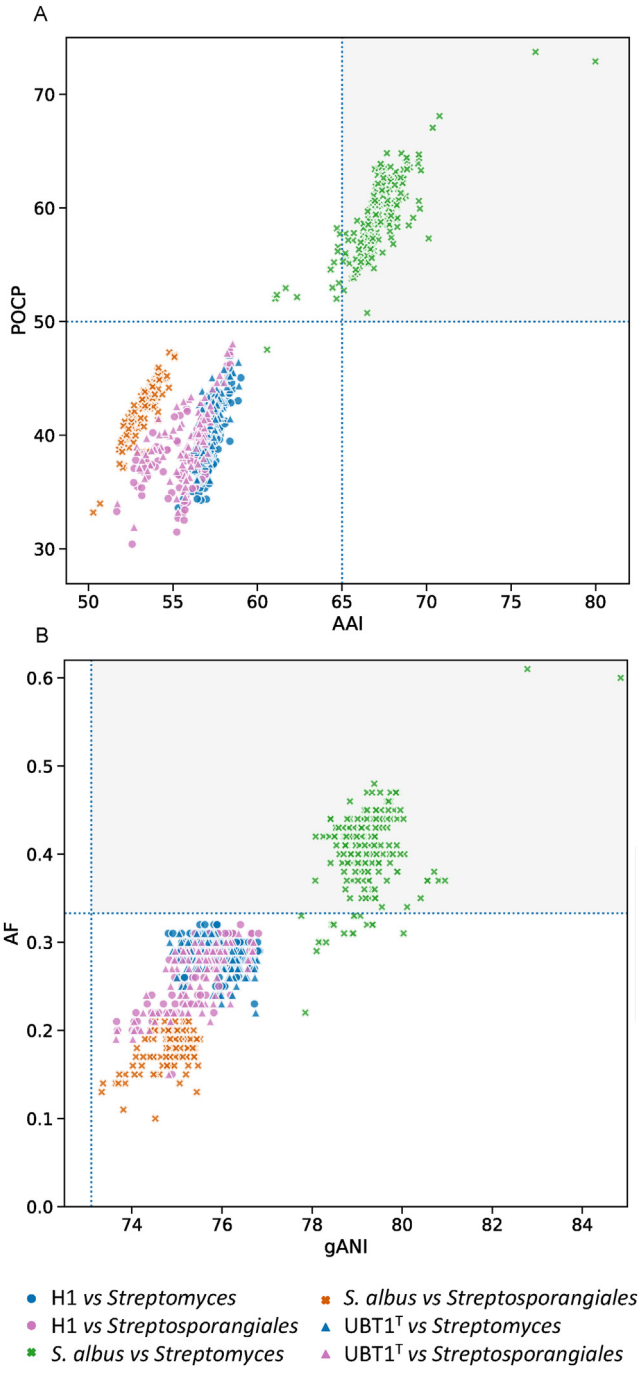

Fig. 4. Distribution of proteomic/genomic metrics between '*S. thermoautotrophicus*' and *Streptosporangiales*/*Streptomyces* type strains. (A) AAI *vs* POCP. (B) gANI *vs* AF. Symbols representing each comparison are depicted in the legend box. Dashed lines represent genus circumscription thresholds. The gray box represents the region where comparisons present proteomic/genomic values that fall inside the genus limits.

phylogenetic approaches measure distances based on variation in the nucleotide or amino acid sequences at each site, or the presence or absence of *indels*, upon an implicit or explicit mathematical model describing the evolution, namely, Bayesian and ML approaches [66]. As exemplified here, the occurrence of incongruence among different tree reconstruction methods are well-known [67,68]. However, we note that the Genome Taxonomy Database (GTDB, release 06-RS202) tree places '*S. thermoautotrophicus*' in the order *Streptomycetales*, thus being congruent with the GBDP tree (Fig. S4). The GTDB approach is based on genome trees inferred with FastTree from an aligned concatenated set of up to 120 single copy marker proteins tree [69,70].

According to the phylogenies demonstrated here, strains UBT1^T and H1 have a distinct phylogenetic position within the class *Actinomycetia*, clearly belonging to a novel family. However, further studies are needed to resolve the ambiguity over the placement of the family, which may represent a novel order.

Genus delineation for UBT1^T and H1 using genomic and proteomic metrics

Despite the advancements in resolving species delineation and the use of genome data to reconstruct the phylogenetic relationship of microorganisms, there is no consensus on the incorporation of genomic metrics and cutoffs to demarcate genera and higher taxa. Nevertheless, different metrics that measure proteomic and genomic relatedness to demarcate genera have been proposed on the basis of AAI [16] and POCP [17]. Recently, Barco et al. [18] utilized the MiSI method [47] for genus delineation, and they verified that the gANI and AF mean values for genus inflection points in Bacteria are 73.1% and 0.333, respectively. Thus, we have applied these approaches to evaluate '*S. thermoautotrophicus*' UBT1^T and H1, *Streptomyces*, *Acidothermus*, and *Streptosporangiales* genomes in detail within the taxonomic context of genus.

In the comparison of the closely related *Actinomycetia* to UBT1^T and H1, different *Streptosporangiales* genomes presented the highest POCP values while some *Streptomyces* genomes present the highest AAI values (Fig. 4A). According to the AAI measure, *Streptomyces megasporus* NRRL B-16372^T is a closely related strain to H1 with 59.0% AAI and 45.1% POCP values, while *Streptomyces vitaminophilus* ATCC 31673^T is closely related to UBT1^T, presenting 58.9% AAI and 44.4% POCP. According to the POCP metric, *Thermomonospora catenispora* 3–22-3^T (*Streptosporangiales*) is closely related to both H1 and UBT1^T presenting 46.2 and 48.0% POCP, and 58.3 and 58.5% AAI, respectively. The comparisons of '*S. thermoautotrophicus*' with *Acidothermus* presented even lower values of ~57.3% AAI and 39.3% POCP. Nevertheless, none of the obtained values surpassed the recommended 65 to 72% [16] and 50% [17] thresholds for the delineation of genera using AAI and POCP metrics, respectively. As expected, *S. albus* was unambiguously grouped with *Streptomyces sensu stricto*, while the comparisons of UBT1^T and H1 strains to *Streptosporangiales* and *Streptomyces* appeared to be distinct from *S. albus* vs *Streptosporangiales*.

Given the proteomic similarity of some *Streptomyces* and *Streptosporangiales* genomes to '*S. thermoautotrophicus*', we further explored the proteomic similarity between UBT1^T and H1 to 102 *Streptosporangiales* and 223 *Streptomyces* genomes. Comparing *Streptomyces* species to UBT1^T and H1, respectively, we found an AAI of 57.20 ± 0.49 (% mean ± SD) and 57.15 ± 0.5, and the number of common proteins to be 2471 ± 116 and 2315 ± 104. For *Streptosporangiales*, we found an AAI value of 55.8 ± 1.5 and 55.6 ± 1.5, and the number of common proteins to be 2403 ± 204 and 2255 ± 184. In this analysis, we also did not find any AAI values $\geq$ 65% to the '*S. thermoautotrophicus*' strains.

In the gANI(AF) analysis (Fig. 4B), similarly to the POCP *vs* AAI correlation plot, *S. albus* was grouped with *Streptomyces* as expected, while the comparisons of UBT1^T and H1 to *Streptosporangiales* were intermixed, and the two strains are clearly distinct from *Streptomyces*. Although some type species from *Streptosporangium*, *Catenulispora*, *Frankia*, *Micromonospora*, *Pseudonocardia*, *Sporichthya* and *Streptomyces* present gANI values that surpass 73.1% in relation to UBT1^T and H1, these comparisons do not surpass the minimum AF requirement for genus definition i.e. gANI and AF are inconsistent. While gANI represents the identity of orthologous genes identified as BBHs using similarity searches, the AF is a complementary measure of the minimum amount that genomes must overlap [47]. If the homologous regions are short with respect to the total length of the genomes, as might be seen following a HGT event, then ANI values may be high even though the bacteria are distantly related. The comparison of UBT1^T and H1 with *A. cellulolyticus* presented 73.0% (~0.15) gANI (AF) (Tables S2 and S3).

The genomic and proteomic metrics results together demonstrated the substantial difference between '*S. thermoautotrophicus*' and other *Actinomycetia* members. Sequences from strains UBT1^T and H1 are clearly below the established cut-off values (gANI-AF: 73.1%-0.333; AAI: 65–72%; POCP: 50%) for defining bacterial genera, strongly suggesting they represent a novel taxon within *Actinomycetia*.

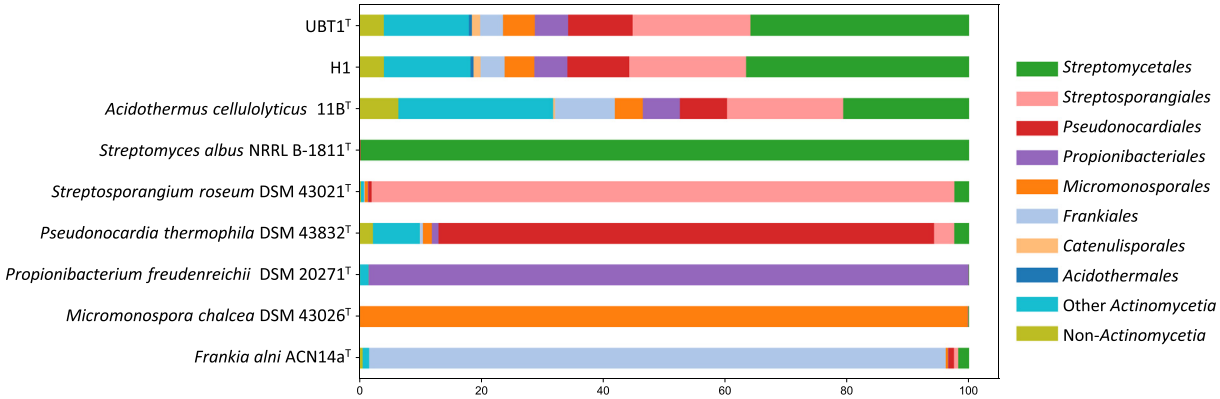

Fig. 5. Taxonomic profile of *Actinomycetia* proteomes. The stacked bar chart shows the proportion of proteins binned according to the taxonomic attribution of the best-hit identified in the database. Taxa are represented by the colored sub-bars.

Taxonomic composition of the 'S. thermoautotrophicus' proteomes

To evaluate the taxonomic composition of the '*S. thermoautotrophicus*' proteomes, we used strain UBT1^T and H1 protein sequences as queries at AAI-profiler for homology searches in the UniProt database. As demonstrated in Fig. 5, *Streptomycetales* proteins were the top hit for only ~36% of the query proteins from strains UBT1^T and H1, while ~19% of them matched to *Streptosporangiales* order proteins. The other query proteins are distributed among different orders of the *Actinomycetia*.

The apparent mosaic nature of the UBT1^T and H1 genomes reflects the underrepresentation of closely related strains in the public sequence databases rather than HGT. Despite the rapid expansion in number of sequenced bacterial and archaeal genomes in the past decade [27,71,72] along with the number of species names validly published [12], understudied groups are often represented by a single family [73–77], along with a few or no genomes present in nucleotide databases. This bias is evident to *A. cellulolyticus*, currently the sole species in *Acidothermus*, the sole genus within *Acidothermaceae*, a unique family within order *Acidothermales* [25]. The query proteins from *A. cellulolyticus*, similarly to UBT1^T and H1, were distributed between many taxonomic groups and there are no *Acidothermales* counterparts in the databases.

According to this analysis, the UBT1^T and H1 proteomes are unique among other members of *Actinomycetia*, corroborating the previous phylogenomic and proteomics/genomics metrics results that indicated a distinctive placement for this taxon.

Phenotypic distinctness of 'S. thermoautotrophicus'

The metabolic distinctiveness of '*S. thermoautotrophicus*' UBT1^T and H1 was predicted based on genome comparisons with 71 *Streptomyces* spp. KO profiles available in the KEGG database. Additional discriminative phenotypic properties were retrieved from the literature for closely related *Actinomycetia* species.

When compared to UBT1^T and H1, 101 KOs were exclusively present among the *Streptomyces* spp. profiles (Table S4). On the other hand, 136 KOs were exclusively present in the UBT1^T and H1 profiles (Table S5), including a nitrate/nitrite sensor two-component system (*narXP*) and multiple genes related to carbon metabolism, such as ribulose-bisphosphate carboxylase (*rbcLS*), glucose/mannose-6-phosphate isomerase, phosphoenolpyruvate carboxykinase, PFK 6-phosphofructokinase 1, fructose 1,6-bisphosphate aldolase/phosphatase, fructose-bisphosphate aldolases, classes I and II. Many exclusive KOs and some Non-Homologous Isofunctional Enzymes (NISEs) cases observed between UBT1^T and H1 and other *Streptomyces* spp. suggest evolutionary divergences in their metabolisms and distant common ancestors. NISEs are evolutionarily unrelated enzymes that catalyze the same biochemical reactions [78]. For example, exclusive KOs for UBT1^T and H1 (K01754) and for other *Streptomyces* spp. (K01752) are related to the same L-serine ⇄ pyruvate + NH_3 enzymatic reaction (R00220) but were exclusively found in each group. While UBT1^T and H1 have some exclusive enzymes, including RuBisCO, related to a carbon autotrophic lifestyle, the other KEGG from *Streptomyces* spp. showed some exclusive KOs related to a heterotrophic lifestyle, including *gluABCD*.

The major characteristic that differentiates UBT1^T from *Acidothermaceae, Nocardiopsaceae, Streptomycetaceae*, *Streptosporangiaceae*, *Thermomonosporaceae*, and *Treboniaceae* is its unique ability to grown chemolithotrophically on CO or CO_2 and H_2 (Table 1). UBT1^T can also be distinguished from these families based on the discontinuous distribution of chemotaxonomic markers, notably cell wall amino acids, menaquinones, and diagnostic sugars in whole cell hydrolysates, in addition to the presence of spores and colony morphology.

Table 1
Comparison of morphological, physiological, and chemotaxonomic characteristics of selected families within *Actinomycetia*.

Characteristic	*Carbonactinosporaceae* (UBT1^T)	*Streptomycetaceae*	*Acidothermaceae*	*Treboniaceae*	*Streptosporangiales*		*Nocardiopsaceae*
					Thermomonosporaceae	*Streptosporangiaceae*	
Morphology	Branched substrate mycelium and scanty aerial mycelium, both with chains of spores[ab]	Substrate mycelium extensively branched, with absent/short chains of spores. Aerial mycelium with chains of few to many spores[c]	Smooth, circular, entire, creamy colonies. Slender rods and filaments, depending on the carbon source provided[hi]	Non-branched substrate mycelium, with single terminal spores. Aerial mycelium absent.[j]	Branched substrate mycelium. Aerial mycelium with single spores/short chains/ spore vesicles with motile spores.[k]	Branched substrate mycelium, with spores when aerial mycelium is absent. Aerial mycelium absent / spore /spore vesicles[k]	Substrate mycelium absent/extensively branched, with single spores/clusters/ spore chains terminating into pseudosporangia. Aerial mycelium with single spores on dichotomously branched sporangia/long or short spore chains.[k]
Spores	Non-motile[a]	Non-motile[d]	Absent[h]	Non-motile[j]	Motile/non-motile[k]	Motile/non-motile[k]	Non-motile[k]
Nutritional type	Facultative chemolithoautotroph[b]	Chemoorganotroph[c]	Prototroph[h]	Chemoorganotroph[j]	Chemoorganotroph[k]	Chemoorganotroph[k]	Chemoorganotroph[k]
Temperature	Thermophilic[a,b]	Mesophilic, thermophilic, psychrophilic[e]	Thermophilic[h]	Mesophilic[j]	Mesophilic, thermophilic[l]	Mesophilic, thermophilic[m]	Mesophilic, thermophilic[op]
Cell-wall predominant diamino acid [†]	LL-DAP[a]	LL- or *meso*-DAP[f]	n/a	*meso*-DAP[j]	*meso*-DAP[k]	*meso*-DAP[k]	*meso*-DAP[k]
Predominant menaquinones [§]	MK-9(H_4)[a]	MK-9(H_6, H_8)[f]	n/a	MK-9(H_6), MK-9 (H_8)[j]	MK-9(H_6)[k]	MK-9(H_2), MK-9(H_4)[k]	MK9, MK10, MK11[k]
Diagnostic sugars [*]	Rib[a]	Gal, Rha, none[f]	n/a	Ara, Gal, Xyl,[j]	Mad, none[k]	Mad, none[k]	Ara, Gal, Rib, none[k]
G + C (%)	69.2–71.0[b]	66–75.3[fg]	66.9[h]	69.6[j]	66–73[k]	64–77[k]	64–76[k]

Data from a: Gadkari et al. [20]; b: MacKellar et al. [19]; c: Kämpfer et al. [79]; d: Nouioui et al. [11]; e: Schrempf [80]; f: Kim et al. [81] and previous studies [82–87]; g: Huang et al. [88]; h: Mohagheghi et al. [89]; i: Berry et al. [90]; j: Rapoport et al. [91]; k: Goodfellow [92]; l: Kroppenstedt and Goodfellow [93]; m: Otoguro et al. [94]; n: Goodfellow et al. [95]; o: Yan et al. [96]; p: Kroppenstedt and Evtushenko [97]. [†] DAP, diaminopimelic acid; [§] MK-9(H_2, H_4, H_6, H_8), di-, tetra-, hexa-, octa- and -hydrogenated menaquinones with nine isoprene units; MK-10 and MK-11 menaquinones with ten and eleven isoprene units; [*] Ara, arabinose; Gal, galactose; Mad, madurose; Rib, ribose; Rha, rhamnose; Xyl, xylose. n/a: data not available.

Table 2
Description of *Carbonactinospora thermoautotrophica* gen. nov., comb. nov.

Genus name	*Carbonactinospora*	–
Species name	–	*Carbonactinospora thermoautotrophica*
Genus status	gen. nov.	–
Genus etymology	Car.bon.ac.ti.no.spo'ra. L. masc. n. *carbo* charcoal; Gr. fem. n. *actis*, *actinos* a ray; Gr. fem. n. *spora* a seed and, in biology, a spore; N.L. fem. n. *Carbonactinospora*, an actinomycete found near a charcoal burning pile	–
Type species of the genus	*Carbonactinospora thermoautotrophica*	–
Specific epithet	–	*thermoautotrophica*
Species status	–	comb. nov.
Species etymology	–	ther.mo.au.to.tro'phi.ca. Gr. masc. adj. *thermos*, hot; Gr. pref. *autos*, self; Gr. masc. adj. *throphikos*, nursing, tending or feeding; N.L. fem. adj. *thermoautotrophica*, heat-loving self-nourishing, referring to the ability to grow chemolithotrophically at high temperature. Basonym: *Streptomyces thermoautotrophicus* Gadkari et al. 1991.
Description of the new taxon and diagnostic traits	Gram-stain positive thermophilic bacteria that form a non-fragmenting, branched substrate mycelium, and aerial hyphae that septate into chains of two to eight oval grey-pigmented spores. Endospores, synnemata, sporangia, or sclerotia are not formed. Non-motile, aerobic. The major menaquinone is MK-$_9$(H_4). Capable of growing chemolithotrophically on CO or CO_2 and H_2. Pyruvate can sustain heterotrophic growth. The predominant cell wall diamino acid is LL-diaminopimelic acid. The whole cell sugar profile contains ribose as a diagnostic sugar. The major polar lipids are phosphatidylethanolamine, diphosphatidylglycerol, phosphatidylglycerol, phosphatidylinositol and phosphatidylinositol mannosides. The predominant fatty acids are iso-$C_{17:0}$, anteiso-$C_{17:0}$, and 14-methylheptadecanoic acid.	The description of the type strain is as given by Gadkari et al. (1990) and MacKellar et al. (2016).
Country of origin	–	Germany
Region of origin	–	Franconian Mountains
Date of isolation	–	1989
Source of isolation	–	Covering soil taken from burning charcoal piles
Genome accession number	–	RefSeq = JYIK01000000
Genome status	–	Complete
Genome size	–	5134 kbp
GC mol%	–	71
Number of strains in study	–	2
Source of isolation of non-type strains	–	Soil taken from a charcoal pile in Hasselfelde, Germany
Information related to the Nagoya Protocol	–	There are no known Nagoya Protocol restrictions for the strains
Designation of the Type Strain	–	UBT1^T
Strain Collection Numbers	–	DSM 100163^T = KCTC 49540^T

Conclusions

Based on the genetic and phenotypic distinctness presented above, we conclude that the chemolithotrophic strains '*S. thermoautotrophicus*' UBT1^T and H1 represent a novel genus, consistent with previous observations [11,19], and for which we propose the name *Carbonactinospora thermoautotrophica* gen. nov., comb. nov. (Table 2). Our additional phylogenomic analysis indicate that the genus *Carbonactinospora* should be placed in a novel family, *Carbonactinosporaceae* fam. nov. In accordance with the current GTDB taxonomy (Fig. S4), the family *Carbonactinosporaceae* is placed within the order *Streptomycetales*, but we note that there are ambiguities in phylogenomic analyses (Figs. 2 and 3) that warrant further studies.

Description of *Carbonactinosporaceae* fam. nov.

(Car.bon.ac.ti.no.spo.ra.ce'ae. N.L. fem. n. *Carbonactinospora*, type genus of the family; *-aceae*, ending to denote a family; N.L. fem. pl. n. *Carbonactinosporaceae*, the *Carbonactinospora* family).

Gram-stain positive, mycelium-forming sporulating bacteria. *Carbonactinosporaceae* represents a distinct *Actinomycetia* phylogenetic lineage based on multigene-based phylogenetic analyses. The type genus is *Carbonactinospora*.

Funding

This work was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brazil), the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil) and KAUST, Saudi Arabia (BAS/1/1096-01-01).

Acknowledgement

We thank Dr Imen Nouioui (Leibniz Institute DSMZ, Germany) for arranging deposit of strain UBT1T (=DSM 100163^T) in the Korean Collection of Type Cultures. We thank Aharon Oren (Hebrew University of Jerusalem, Israel) for his advice on the formation of the Latin names proposed.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.syapm.2021.126223.

References

[1] Salam, N., Jiao, J.Y., Zhang, X.T., Li, W.J. (2020) Update on the classification of higher ranks in the phylum Actinobacteria. Int. J. Syst. Evol. Microbiol. 70, 1331–55.

[2] Chater, K.F. (2016) Recent advances in understanding *Streptomyces*. F1000Res. 5, 2795.

[3] Hopwood, D.A. (2019) Highlights of *Streptomyces* genetics. Heredity 123 (1), 23–32.

[4] Nodwell, J.R. (2019) Microbe Profile: *Streptomyces coelicolor*: a burlesque of pigments and phenotypes. Microbiology 165, 953–955.

[5] Belknap, K.C., Park, C.J., Barth, B.M., Andam, C.P. (2020) Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. Sci. Rep. 10, 2003.

[6] Chevrette, M.G., Carlos-Shanley, C., Louie, K.B., Bowen, B.P., Northen, T.R., Currie, C.R. (2019) Taxonomic and metabolic incongruence in the ancient genus *Streptomyces*. Front. Microbiol. 10, 2170.

[7] van Bergeijk, D.A., Terlouw, B.R., Medema, M.H., van Wezel, G.P. (2020) Ecology and genomics of *Actinobacteria*: new concepts for natural product discovery. Nat. Rev. Microbiol. 18 (10), 546–558.

[8] Labeda, D.P., Goodfellow, M., Brown, R., Ward, A.C., Lanoot, B., Vanncanneyt, M., Swings, J., Kim, S.-B., Liu, Z., Chun, J., Tamura, T., Oguchi, A., Kikuchi, T., Kikuchi, H., Nishii, T., Tsuji, K., Yamaguchi, Y., Tase, A., Takahashi, M., Sakane, T., Suzuki, K.I., Hatano, K. (2012) Phylogenetic study of the species within the family *Streptomycetaceae*. Antonie Van Leeuwenhoek 101 (1), 73–104.

[9] Cheng, K., Rong, X., Huang, Y. (2016) Widespread interspecies homologous recombination reveals reticulate evolution within the genus *Streptomyces*. Mol. Phylogenet. Evol. 102, 246–254.

[10] Labeda, D.P., Dunlap, C.A., Rong, X., Huang, Y., Doroghazi, J.R., Ju, K.-S., Metcalf, W.W. (2017) Phylogenetic relationships in the family *Streptomycetaceae* using multi-locus sequence analysis. Antonie Van Leeuwenhoek 110 (4), 563–583.

[11] Nouioui, I., Carro, L., García-López, M., Meier-Kolthoff, J.P., Woyke, T., Kyrpides, N.C., Pukall, R., Klenk, H.P., Goodfellow, M., Göker, M. (2018) Genome-based taxonomic classification of the phylum *Actinobacteria*. Front. Microbiol. 9, 2007.

[12] Sant'Anna, F.H., Bach, E., Porto, R.Z., Guella, F., Sant'Anna, E.H., Passaglia, L.M.P. (2019) Genomic metrics made easy: what to do and where to go in the new era of bacterial taxonomy. Crit. Rev. Microbiol. 45, 182–200.

[13] Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D.R., da Costa, M.S., Rooney, A.P., Yi, H., Xu, X.W., De Meyer, S., Trujillo, M.E. (2018) Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int. J. Syst. Evol. Microbiol. 68, 461–466.

[14] Konstantinidis, K.T., Tiedje, J.M. (2005) Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A. 102, 2567–2572.

[15] Richter, M., Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. Proc. Natl. Acad. Sci. U. S. A. 106, 19126–19131.

[16] Konstantinidis, K.T., Tiedje, J.M. (2007) Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. Curr. Opin. Microbiol. 10 (5), 504–509.

[17] Qin, Q.-L., Xie, B.-B., Zhang, X.-Y., Chen, X.-L., Zhou, B.-C., Zhou, J., Oren, A., Zhang, Y.-Z. (2014) A proposed genus boundary for the prokaryotes based on genomic insights. J. Bacteriol. 196 (12), 2210–2215.

[18] Barco, R.A., Garrity, G.M., Scott, J.J., Amend, J.P., Nealson, K.H., Emerson, D., Giovannoni, S.J. (2020) A genus definition for *Bacteria* and *Archaea* based on a standard genome relatedness index. MBio 11 (1). https://doi.org/10.1128/mBio.02475-19.

[19] MacKellar, D., Lieber, L., Norman, J.S., Bolger, A., Tobin, C., Murray, J.W., Oksaksin, M., Chang, R.L., Ford, T.J., Nguyen, P.Q., Woodward, J. (2016) *Streptomyces thermoautotrophicus* does not fix nitrogen. Sci. Rep. 6, 20086.

[20] Gadkari, D., Schricker, K., Acker, G., Kroppenstedt, R.M., Meyer, O. (1990) *Streptomyces thermoautotrophicus* sp. nov., a thermophilic CO- and H_2-oxidizing obligate chemolithoautotroph. Appl. Environ. Microbiol. 56, 3727–3734.

[21] Ribbe, M., Gadkari, D., Meyer, O. (1997) N_2 fixation by *Streptomyces thermoautotrophicus* involves a molybdenum-dinitrogenase and a manganese-superoxide oxidoreductase that couple N_2 reduction to the oxidation of superoxide produced from O_2 by a molybdenum-CO dehydrogenase. J. Biol. Chem. 272, 26627–26633.

[22] Zhao, Y., Bian, S.M., Zhou, H.N., Huang, J.F. (2006) Diversity of nitrogenase systems in diazotrophs. J. Integr. Plant. Biol. 48, 745–755.

[23] Szafran, M.J., Jakimowicz, D., Elliot, M.A. (2020) Compaction and control-the role of chromosome-organizing proteins in *Streptomyces*. FEMS Microbiol. Rev. 44, 725–739.

[24] Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T., Peplies, Jörg, Ludwig, W., Glöckner, F.O. (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 42 (D1), D643–D648.

[25] Sen, A., Daubin, V., Abrouk, D., Gifford, I., Berry, A.M., Normand, P. (2014) Phylogeny of the class *Actinobacteria* revisited in the light of complete genomes. The orders "*Frankiales*" and *Micrococcales* should be split into coherent entities: proposal of *Frankiales* ord. nov., *Geodermatophilales* ord. nov., *Acidothermales* ord. nov. and *Nakamurellales* ord. nov. Int. J. Syst. Evol. Microbiol. 64, 3821–3832.

[26] Pruesse, E., Peplies, J., Glöckner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28, 1823–1829.

[27] Yoon, S.H., Ha, S.M., Kwon, S., Lim, J., Kim, Y., Seo, H., Chun, J. (2017) Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int. J. Syst. Evol. Microbiol. 67, 1613–1617.

[28] Wu, M., Scott, A.J. (2012) Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinformatics 28, 1033–1034.

[29] Seah, B. (2016) Phylogenomics-tools: first release. Zenodo. https://doi.org/10.5281/zenodo.46122.

[30] Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32 (5), 1792–1797.

[31] Guindon, S., Lethiec, F., Duroux, P., Gascuel, O. (2005) PHYML Online - a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 33 (Web Server), W557–W559.

[32] Anisimova, M., Gil, M., Dufayard, J.F., Dessimoz, C., Gascuel, O. (2011) Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst. Biol. 60, 685–699.

[33] Contreras-Moreira, B., Vinuesa, P. (2013) GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. Appl. Environ. Microbiol. 79, 7696–7701.

[34] Vinuesa, P., Ochoa-Sánchez, L.E., Contreras-Moreira, B. (2018) GET_PHYLOMARKERS, a software package to select optimal orthologous clusters for phylogenomics and inferring pan-genome phylogenies, used for a critical geno-taxonomic revision of the genus. Front. Microbiol. 9, 771.

[35] Junier, T., Zdobnov, E.M. (2010) The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics 26 (13), 1669–1670.

[36] Meier-Kolthoff, J.P., Göker, M. (2019) TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat. Commun. 10, 2182.

[37] Ondov, B.D., Treangen, T.J., Melsted, P., Mallonee, A.B., Bergman, N.H., Koren, S., Phillippy, A.M. (2016) Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 17, 132.

[38] Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.H., Rognes, T., Ussery, D.W. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3100–3108.

[39] Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L. (2009) BLAST+: architecture and applications. BMC Bioinformatics 10 (1), 421. https://doi.org/10.1186/1471-2105-10-421.

[40] Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14 (1), 60. https://doi.org/10.1186/1471-2105-14-60.

[41] Auch, A.F., Klenk, H.P., Göker, M. (2010) Standard operating procedure for calculating genome-to-genome distances based on high-scoring segment pairs. Stand. Genomic. Sci. 2, 142–148.

[42] Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.P., Göker, M. (2014) Highly parallelized inference of large genome-based phylogenies. Concurr. Comput. 26, 1715–1729.

[43] Henz, S.R., Huson, D.H., Auch, A.F., Nieselt-Struwe, K., Schuster, S.C. (2005) Whole-genome prokaryotic phylogeny. Bioinformatics 21 (10), 2329–2335.

[44] Lefort, V., Desper, R., Gascuel, O. (2015) FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. Mol. Biol. Evol. 32, 2798–2800.

[45] Farris, J.S. (1972) Estimating phylogenetic trees from distance matrices. Am. Nat. 106, 645–667.

[46] Kreft, L., Botzki, A., Coppens, F., Vandepoele, K., Van Bel, M. (2017) PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. Bioinformatics 33, 2946–2947.

[47] Varghese, N.J., Mukherjee, S., Ivanova, N., Konstantinidis, K.T., Mavrommatis, K., Kyrpides, N.C., Pati, A. (2015) Microbial species delineation using whole genome sequences. Nucleic Acids Res. 43 (14), 6761–6771.

[48] Rodriguez, R.L.M., Konstantinidis, K.T. (2016) The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ. e1900v1.

[49] Konstantinidis, K.T., Tiedje, J.M. (2005) Towards a genome-based taxonomy for prokaryotes. J. Bacteriol. 187 (18), 6258–6264.

[50] Medlar, A.J., Törönen, P., Holm, L. (2018) AAI-profiler: fast proteome-wide exploratory analysis reveals taxonomic identity, misclassification and contamination. Nucleic Acids Res. 46, W479–485.

[51] Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., Kanehisa, M. (1999) KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 27 (1), 29–34.

[52] Stoddard, S.F., Smith, B.J., Hein, R., Roller, B.R.K., Schmidt, T.M. (2015) rrnDB: improved tools for interpreting rRNA gene abundance in Bacteria and Archaea and a new foundation for future development. Nucleic Acids Res. 43, D593–598.

[53] Sun, D.L., Jiang, X., Wu, Q.L., Zhou, N.Y. (2013) Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. Appl. Environ. Microbiol. 79, 5962–5969.

[54] Acinas, S.G., Marcelino, L.A., Klepac-Ceraj, V., Polz, M.F. (2004) Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. J. Bacteriol. 186 (9), 2629–2635.

[55] Klappenbach, J.A., Dunbar, J.M., Schmidt, T.M. (2000) rRNA operon copy number reflects ecological strategies of *Bacteria*. Appl. Environ. Microbiol. 66, 1328–1333.

[56] Roller, B.R.K., Stoddard, S.F., Schmidt, T.M. (2016) Exploiting rRNA operon copy number to investigate bacterial reproductive strategies. Nat. Microbiol. 1, 16160.
[57] Valdivia-Anistro, J.A., Eguiarte-Fruns, L.E., Delgado-Sapién, G., Márquez-Zacarías, P., Gasca-Pineda, J., Learned, J., Elser, J.J., Olmedo-Alvarez, G., Souza, V. (2015) Variability of rRNA operon copy number and growth rate dynamics of *Bacillus* isolated from an extremely oligotrophic aquatic ecosystem. Front. Microbiol. 6, 1486.
[58] Salwan, R., Sharma, V. Overview of extremophiles. In: Salwan, R., Sharma, V. (Eds.), Physiological and Biotechnological Aspects of Extremophiles. Academic Press, Cambridge, pp. 3–11.
[59] López-López, A., Benlloch, S., Bonfá, M., Rodríguez-Valera, F., Mira, A. (2007) Intragenomic 16S rDNA divergence in *Haloarcula marismortui* is an adaptation to different temperatures. J. Mol. Evol. 65, 687–696.
[60] Suzuki, M.T., Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62, 625–630.
[61] Hong, S., Bunge, J., Leslin, C., Jeon, S., Epstein, S.S. (2009) Polymerase chain reaction primers miss half of rRNA microbial diversity. ISME J. 3 (12), 1365–1373.
[62] Yarza, P., Yilmaz, P., Pruesse, E., Glöckner, F.O., Ludwig, W., Schleifer, K.H., Whitman, W.B., Euzéby, J., Amann, R., Rosselló-Móra, R. (2014) Uniting the classification of cultured and uncultured *Bacteria* and *Archaea* using 16S rRNA gene sequences. Nat. Rev. Microbiol. 12, 635–645.
[63] Lewin, G.R., Carlos, C., Chevrette, M.G., Horn, H.A., McDonald, B.R., Stankey, R.J., Fox, B.G., Currie, C.R. (2016) Evolution and ecology of *Actinobacteria* and their bioenergy applications. Annu. Rev. Microbiol. 70 (1), 235–254.
[64] McDonald, B.R., Currie, C.R., Keim, P., Marx, C., Abbot, P. (2017) Lateral gene transfer dynamics in the ancient bacterial genus *Streptomyces*. MBio 8 (3). https://doi.org/10.1128/mBio.00644-17.
[65] Van de Peer, Y., Salemi, M. Phylogenetic inference based on distance methods. In: Lemey, P., Salemi, M., Vandamme, A.M. (Eds.), The Phylogenetic Handbook. Cambridge University Press, Cambridge, pp. 142–180.
[66] Gil, M., Anisimova, M. Methodologies for phylogenetic inference. Encyclopedia of Life.
[67] Sciences (ELS) subject area: Evolution & Diversity of Life, John Wiley & Sons, Ltd, Chichester, pp. 1–5.
[68] Degnan, J.H., Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340.
[69] Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H. (2006) Phylogenomics: the beginning of incongruence? Trends Genet. 22, 225–231.
[70] Parks, D.H., Chuvochina, M., Chaumeil, P.-A., Rinke, C., Mussig, A.J., Hugenholtz, P. (2020) A complete domain-to-species taxonomy for *Bacteria* and *Archaea*. Nat. Biotechnol. 38 (9), 1079–1086.
[71] Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36 (10), 996–1004.
[72] Garrity, G.M., Kraft, C.S. (2016) A new genomics-driven taxonomy of *Bacteria* and *Archaea*: Are we there yet?. J. Clin. Microbiol. 54 (8), 1956–1963.
[73] Hugenholtz, P., Skarshewski, A., Parks, D.H. (2016) Genome-based microbial taxonomy coming of age. Cold Spring Harb. Perspect. Biol. 8 (6), a018085. https://doi.org/10.1101/cshperspect.a018085.
[74] Tahon, G., Tytgat, B., Lebbe, L., Carlier, A., Willems, A. (2018) *Abditibacterium utsteinense* sp. nov., the first cultivated member of candidate phylum FBP, isolated from ice-free Antarctic soil samples. Syst. Appl. Microbiol. 41, 279–290.
[75] Qin, W., Heal, K.R., Ramdasi, R., Kobelt, J.N., Martens-Habbena, W., Bertagnolli, A.D., Amin, S.A., Walker, C.B., Urakawa, H., Könneke, M., Devol, A.H. (2017) Nitrosopumilus maritimus gen. nov., sp. nov., Nitrosopumilus cobalaminigenes sp. nov., Nitrosopumilus oxyclinae sp. nov., and Nitrosopumilus ureiphilus sp. nov., four marine ammonia-oxidizing archaea of the phylum Thaumarchaeota. Int. J. Syst. Evol. Microbiol. 67, 5067–5079.
[76] Jumas-Bilak, E., Marchandin, H., Roudière, L. (2009) Description of 'Synergistetes' phyl. nov. and emended description of the phylum 'Deferribacteres' and of the family Syntrophomonadaceae, phylum 'Firmicutes.' Int. J. Syst. Evol. Microbiol. 59, 1851–1851.
[77] Garrity, G.M., Bell, J.A., Lilburn, T. (2005) Acidithiobacillales ord. nov. In: Brenner, D.J. et al. (Eds.), Bergey's Manual® of Systematic Bacteriology, Springer, Boston, pp. 60–63.
[78] Omelchenko, M.V., Galperin, M.Y., Wolf, Y.I., Koonin, E.V. (2010) Non-homologous isofunctional enzymes: A systematic analysis of alternative solutions in enzyme evolution. Biol. Direct 5 (1), 31. https://doi.org/10.1186/1745-6150-5-31.
[79] Kämpfer, P., Glaeser, S.P., Parkes, L., van Keulen, G., Dyson, P.T., Streptomycetaceae, F.. In: The Prokaryotes. Springer, Berlin, pp. 889–1010.
[80] Schrempf, H. *Streptomycetaceae*: life style, genome, metabolism and habitats. In: Encyclopedia of Life Sciences (ELS) subject area: Microbiology. John Wiley & Sons Ltd, Chichester, pp. 1–7.
[81] Kim, S.B., Lonsdale, J., Seong, C.N., Goodfellow, M. (2003) Streptacidiphilus gen. nov., acidophilic actinomycetes with wall chemotype I and emendation of the family Streptomycetaceae (Waksman and Henrici (1943)AL) emend. Rainey et al. 1997. Antonie Van Leeuwenhoek 83, 107–116.
[82] Shirling, E.B., Gottlieb, D. Retrospective evaluation of International *Streptomyces* Project taxonomic criteria. In: Arai, T. (Ed.), *Actinomycetes*: The Boundary Microorganisms. University Park Press, Baltimore, pp. 9–41.
[83] Ōmura, S., Takahashi, Y., Iwai, Y. (1989) Genus Kitasatosporia Ōmura et al. (1983), 672VP. In: Williams, S.T., Sharpe, M.E., Holt, J.G., (Eds.) Bergey's Manual® of Systematic Bacteriology, Williams & Wilkins, Baltimore, pp. 2594–2598.
[84] Lonsdale, J.T. (1985) Aspects of the biology of acidophilic actinomycetes. Ph.D. Thesis, University of Newcastle, Newcastle, United Kingdom.
[85] Williams, S.T., Goodfellow, M., Alderson, G. (1989) Genus Streptomyces Waksman and Henrici (1943), 339AL. In: Williams, S.T., Sharpe, M.E., Holt, J. G., (Eds.), Bergey's Manual® of Systematic Bacteriology, Williams & Wilkins, Baltimore, pp. 2594–2598.
[86] Nakagaito, Y., Yokota, A., Hasegawa, T. (1992) Three new species of the genus *Streptomyces*: *Streptomyces cochleatus* sp. nov., *Streptomyces paracochleatus* sp. nov., and *Streptomyces azaticus* sp. nov. J. Gen. Appl. Microbiol. 38 (2), 105–120.
[87] Labeda, D.P. (1998) *Kitasatosporia mediocidica* sp. nov. Int. J. Syst. Bacteriol. 38, 287–290.
[88] Huang, M.J., Rao, M.P.N., Salam, N., Xiao, M., Huang, H.Q., Li, W.J. (2017) *Allostreptomyces psammosilenae* gen. nov., sp. nov., an endophytic actinobacterium isolated from the roots of *Psammosilene tunicoides* and emended description of the family *Streptomycetaceae* [Waksman and Henrici (1943) AL] emend. Rainey et al. 1997, emend. Kim et al. 2003, emend. Zhi et al. 2009. Int. J. Syst. Evol. Microbiol. 67, 288–293.
[89] Mohagheghi, A., Grohmann, K., Himmel, M., Leighton, L., Updegraff, D.M. (1986) Isolation and characterization of *Acidothermus cellulolyticus* gen. nov., sp. nov., a new genus of thermophilic, acidophilic, cellulolytic bacteria. Int. J. Syst. Bacteriol. 36, 435–443.
[90] Berry, A.M., Barabote, R.D., Normand, P. The family *Acidothermaceae*. In: Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F. (Eds.), The Prokaryotes. Springer, Berlin, pp. 13–19.
[91] Rapoport, D., Sagova-Mareckova, M., Sedláček, I., Provaznik, J., Králová, S., Pavlinic, D., Benes, V., Kopecky, J. (2020) *Trebonia kvetii* gen. nov., sp. nov., an acidophilic actinobacterium, and proposal of the new actinobacterial family *Treboniaceae* fam. nov. Int. J. Syst. Evol. Microbiol. 70, 5106–5114.
[92] Goodfellow, M. Order XV. *Streptosporangiales* ord. nov. In: Parte, A., Whitman, W.B., Goodfellow, M., Kämpfer, P., Busse, H.J., Trujillo, M.E., Ludwig, W. (Eds.), Bergey's Manual® of Systematic Bacteriology. Springer, New York, pp. 1805–1966.
[93] Kroppenstedt, R.M., Goodfellow, M. The Family *Thermomonosporaceae: Actinocorallia, Actinomadura, Spirillospora* and *Thermomonospora*. In: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E. (Eds.), The Prokaryotes. Springer New York, New York, NY, pp. 682–724. https://doi.org/10.1007/0-387-30743-5_27.
[94] Otoguro, M., Yamamura, H., Quintana, E.T. The family *Streptosporangiaceae*. In: Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F. (Eds.), The Prokaryotes. Springer, Berlin, pp. 1011–1045.
[95] Goodfellow, M. The Family *Nocardiaceae*. In: Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., Thompson, F. (Eds.), The Prokaryotes. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 595–650. https://doi.org/10.1007/978-3-642-30138-4_404.
[96] Yan, X., Yan, H., Liu, Z., Liu, X., Mo, H., Zhang, L. (2011) *Nocardiopsis yanglingensis* sp. nov., a thermophilic strain isolated from a compost of button mushrooms. Antonie Van Leeuwenhoek 100 (3), 415–419.
[97] Kroppenstedt, R.M., Evtushenko, L.I. The Family *Nocardiopsaceae*. In: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E. (Eds.), The Prokaryotes. Springer New York, New York, NY, pp. 754–795.